
Question-Réponse multilingue : évaluation et amélioration des stratégies de changement de langue

Anne-Laure Ligozat — Brigitte Grau
Isabelle Robba — Anne Vilnat

LIMSI-CNRS
BP 133
91403 ORSAY Cedex
prenom.nom@limsi.fr

RÉSUMÉ. Cet article présente un système de question-réponse bilingue, capable de traiter des questions en français en cherchant la réponse dans des documents en anglais (ou potentiellement l'inverse). Deux stratégies de passage d'une langue à l'autre y sont décrites et évaluées. Ces stratégies concernent à la fois la recherche d'information dans le corpus et l'extraction de la réponse. Dans un premier temps, nous étudions l'apport de la traduction de bitermes, et l'influence de la complétion des dictionnaires de traduction. Dans un second temps, nous proposons une stratégie de transfert de l'analyse des questions d'une langue à l'autre et évaluons son influence dans notre chaîne.

ABSTRACT. This article presents a bilingual question answering system, which is able to process questions in French and to find the answer in documents in English (or eventually the other way around). Two cross-lingual strategies are described and evaluated, which concern at least the information retrieval steps, but can also be used by the answer extraction module. First, we study the contribution of biterms translation, and the influence of the completion of the translation dictionaries. Then, we propose a strategy for transferring the question analysis from one language to the other, and we study its influence on the performance of our system.

MOTS-CLÉS : Question-Réponse, multilinguisme.

KEYWORDS: Question answering, multilingualism.

1. Introduction

Lorsqu'une question est posée dans une langue sur le Web, il peut être intéressant de rechercher la réponse à cette question dans des documents écrits dans d'autres langues afin d'augmenter le nombre de documents retournés. C'est ce type de possibilités dont l'évaluation de systèmes de question-réponse multilingues ¹ CLEF souhaite encourager le développement.

L'objectif des systèmes de question-réponse est de retourner des réponses précises à des questions en langage naturel, au lieu de la liste de documents habituellement retournée par un moteur de recherche. L'ouverture sur le multilinguisme des systèmes de question-réponse pose des problématiques qui recoupent en partie celles de la Recherche d'Information multilingue, mais doivent également tenir compte des besoins de l'extraction d'information permettant de trouver la réponse précise. Cet article présente un système de question-réponse bilingue, capable de traiter des questions et des documents en français et en anglais. L'objectif est d'évaluer deux stratégies de passage d'une langue à l'autre, ainsi que leur combinaison afin de dégager plusieurs possibilités d'évolutions.

2. Présentation du système de question-réponse

Notre système de question-réponse bilingue a participé à la campagne d'évaluation CLEF 2005 ², et notamment à la tâche pour laquelle les questions sont en français et les documents à traiter en anglais ³. Ce système est composé de plusieurs modules présentés figure 1 et dont une description plus détaillée pourra être trouvée dans [FER 02]. Le premier module est celui de l'analyse des questions, dont le but est de détecter un certain nombre de leurs caractéristiques qui permettront de trouver les réponses dans les documents. Puis la collection est parcourue grâce au moteur de recherche MG ⁴. Les documents retournés sont ensuite réindexés en fonction de la présence des termes de la question ou de leurs variantes, et plus précisément de leur nombre et de leur type ; puis un module permet de reconnaître les entités nommées, et les phrases des documents sélectionnés sont pondérées en fonction des informations sur la question. Enfin, différents processus dépendant du type attendu de la réponse sont appliqués dans le but d'extraire les réponses des phrases.

1. Nous parlerons dans cet article de bilinguisme ou de multilinguisme pour les systèmes de question-réponse permettant de poser une question dans une langue donnée et d'obtenir une réponse dans une autre langue donnée, car ces dénominations sont couramment adoptées en français. Il s'agit en réalité de systèmes plus bilingues ou translangues ou interlangues (*cross-language* en anglais) que réellement multilingues, le terme de multilinguisme désignant plutôt des systèmes répondant à une question dans une langue donnée en différentes langues.

2. Multilingual Question Answering task at the Cross Language Evaluation Forum, <http://clef-qa.itc.it/>

3. Nous ne nous intéresserons qu'à ce sens dans la suite de l'article, puisque c'est celui pour lequel nous disposons d'évaluations officielles.

4. MG for Managing Gigabytes <http://www.cs.mu.oz.au/mg/>

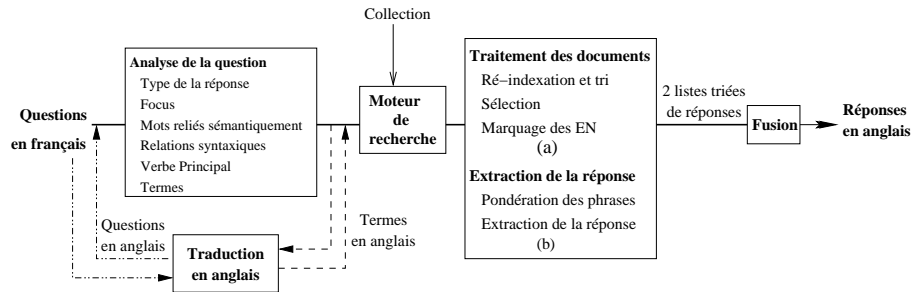


Figure 1. Architecture du système de question-réponse multilingue

3. Gestion du bilinguisme en question-réponse

Deux approches principales sont possibles pour gérer le multilinguisme dans les systèmes de question-réponse : la traduction de la question et la traduction terme à terme, après analyse de la question. Ces deux approches ont été utilisées par les participants aux évaluations CLEF. La première approche de traduction automatique de la question est largement répandue, et a été choisie par exemple par les systèmes suivants : [PER 04], [JIJ 04], [NEU 05], [PAB 05], [TAN 05]. Parmi ces systèmes, plusieurs ont calculé la perte de performances entre système monolingue et multilingue. Ainsi, la version anglais-néerlandais de [JIJ 04] perd 10,5 % de bonnes réponses, passant de 45,5% à 35% de bonnes réponses. La version anglais-français du système de [PER 04] perd 11% de bonnes réponses : le nombre de bonnes réponses passe de 24,5% à 13,5%. [PAB 05] perdent quant à eux entre 5 et 7% de bonnes réponses en passant d'un système monolingue espagnol à un système bilingue anglais-espagnol. [HAR 05] a également mené une expérience avec traduction des questions de l'anglais à l'allemand, et rapporte une chute d'environ 50% des performances de son système. Pour leur système multilingue, [NEU 04] ont choisi d'utiliser plusieurs traducteurs automatiques, et de rassembler les différentes traductions dans un "sac de mots" qui est utilisé pour étendre les requêtes. Des synonymes sont également ajoutés au "sac de mots" et EuroWordNet⁵ est utilisé pour l'étape de désambiguïsation. Cette technique leur permet d'avoir des performances très proches en monolingue et en bilingue, puisque leur système monolingue anglais obtient 25% de bonnes réponses, et leur système bilingue allemand-anglais 23,5%. La traduction automatique de la question pose deux problèmes principaux : elle peut produire des questions syntaxiquement incorrectes, et apporter une mauvaise résolution des ambiguïtés de traduction. De plus, les mots inconnus comme certains noms propres ne sont pas ou mal traduits alors qu'ils devraient l'être (pour la question "Qui est Djokhar Doudayev?", le nom propre n'est pas traduit par son équivalent anglais "Dzhokhar Dudayev", mais est laissé inchangé).

5. Base de données lexicale multilingue comprenant des WordNet pour plusieurs langues européennes, <http://www.ilc.uva.nl/EuroWordNet/>

Nous décrivons plus loin plusieurs possibilités pour tenter de contourner le premier problème, ainsi que notre propre solution.

D'autres systèmes comme [SUT 05] ou [TAN 04] utilisent une traduction terme à terme. [TAN 04], qui ont participé aux tâches bulgare-anglais et italien-anglais en 2004, traduisent les mots-clefs de la question en utilisant des dictionnaires bilingues et MultiWordNet⁶. Afin de diminuer le bruit apporté par les différentes traductions et d'obtenir une meilleure cohésion, ils valident les traductions dans deux vastes corpus, AQUAINT et TIPSTER. Ce système a eu en 2004 un score de 22,5% réponses correctes en bilingue, contre 28% en monolingue. [SUT 05] combinent quant à eux deux systèmes de traduction automatiques et un dictionnaire pour traduire des expressions. Enfin, [LAU 05] traduisent uniquement des mots ou des idiomes, en utilisant l'anglais comme langue pivot. Les performances de ce système passent de 64% de bonnes réponses en monolingue français-français à 39,5% en bilingue anglais-français.

4. Approche adoptée

Afin de gérer le passage du français à l'anglais dans notre système, les stratégies présentées ci-dessus sont appliquées en parallèle : un premier sous-système appelé MUSQUAT traduit les termes issus de l'analyse des questions, et recherche les réponses dans la collection à partir de ces termes traduits ; un second sous-système utilise les traductions des questions par un système de traduction automatique puis le système monolingue anglais QALC est appliqué. Si elles représentent les stratégies les plus courantes pour ce type de tâche, une originalité de notre système est de les implémenter conjointement, ce qui nous permet de fusionner les résultats pour améliorer nos performances. Pour analyser celles-ci, nous présentons les résultats obtenus au cours de la campagne d'évaluation CLEF 2005 dans le tableau 1. Ces chiffres représentent les résultats obtenus à deux étapes différentes de la chaîne de question-réponse. La première ligne indique le pourcentage de questions (sur les 200 constituant l'évaluation) pour lesquelles la bonne réponse est présente dans l'une des cinq premières phrases sélectionnées, après la sélection des phrases opérée au point (a) de la figure 1. Les lignes suivantes indiquent le pourcentage de questions pour lesquelles une réponse est trouvée à la fois au premier rang et dans les cinq premiers rangs, après l'extraction de la réponse (à savoir au point (b) de la figure 1), en distinguant les réponses de type entité nommée (lieu, date, personne...), car la stratégie suivie est différente des autres questions. Enfin la dernière ligne donne les résultats obtenus après fusion des deux stratégies, en tenant compte des cinq premières réponses trouvées par les deux stratégies précédentes. Ce score est celui que nous avons obtenu lors de la campagne CLEF 2005, nous situant juste devant [SUT 05] sur le même couple de langue, et troisième des équipes ayant l'anglais comme langue cible (le meilleur score est obtenu avec un passage allemand vers anglais avec 25,5%). L'ensemble des résultats est disponible dans [VAL 05].

6. Base de données lexicale multilingue dans laquelle le WordNet italien est aligné avec le WordNet de Princeton, <http://multiwordnet.itc.it>

		MUSQUAT %	QALC %
Phrases	5 premiers rangs	41	46
Réponses correctes EN	rang 1	18	14
	5 premiers rangs	26	17
Réponses correctes nonEN	rang 1	16	13
	5 premiers rangs	23	20
Total	rang 1	17	13
	5 premiers rangs	24	19
Fusion des deux stratégies		19	

Tableau 1. Performances de notre système à CLEF 2005

Comme le montre ce tableau, les deux stratégies de gestion du bilinguisme présentent des résultats très différents, ce que les caractéristiques de chaque stratégie permettent d'expliquer. En effet, la traduction des termes de la question effectuée par MUSQUAT fournit au module d'analyse des questions, des questions en français bien formulées, dont l'analyse sera par conséquent plus fiable. En revanche, la traduction des termes se fait alors par recherche de toutes les traductions possibles pour chaque terme, et donc s'effectue hors contexte, dépend de la qualité des dictionnaires, et introduit du bruit du fait de l'utilisation de traductions erronées. Dans MUSQUAT, les traductions de termes ne se limitent cependant pas aux termes constitués d'un seul mot : les bitermes (soit les termes constitués de deux mots liés) de la question française sont également recherchés. Ces bitermes sont construits à partir de l'analyse de la question : toute suite de 2 termes du type *adjectif/nom commun* ou encore *nom propre/nom propre...* forme un biterme. Ils sont traduits, en combinant toutes les traductions possibles, puis ces différentes combinaisons sont recherchées dans le corpus, dans le but de valider celles qui y sont trouvées. Les bitermes anglais ainsi obtenus sont ensuite utilisés par les autres modules du système. L'intérêt d'utiliser les bitermes vient du fait que ceux-ci permettent un début de désambiguïsation en donnant un contexte minimal aux mots qui les composent. De plus, notre hypothèse suite aux résultats obtenus lors de la campagne de CLEF 2004 [GRA 05a] était que la présence des traductions des bitermes dans le corpus confirme les traductions des monoterms, ce que nous avons pu vérifier dans [GRA 05b], où cette stratégie est décrite en détail.

La traduction de la question avant lancement de QALC présente quant à elle l'avantage de fournir une traduction unique des termes, qui se révèle généralement assez fiable. Cependant, la grammaticalité ou le réalisme de la question ne sont pas assurés, et par conséquent son analyse, fondée sur des patrons, est perturbée.

Dans le travail présenté ici, nous avons cherché à évaluer et contourner les difficultés posées par chacune des stratégies : d'une part (paragraphe 5), en examinant comment la traduction des bitermes dans MUSQUAT pourrait être rendue plus fiable, et d'autre part (paragraphe 6), en améliorant l'analyse des questions, en s'appuyant sur les questions en français, pour QALC.

5. Traduction des bitermes

La traduction des termes et des bitermes présents dans les questions est effectuée dans MUSQAT à l'aide de deux dictionnaires. Le premier, utilisé seul lors de notre première participation à CLEF l'an dernier est Magic-Dic ⁷. Dictionnaire sous licence GPL, il a été retenu pour sa capacité à évoluer : les utilisateurs peuvent proposer de nouvelles traductions qui sont vérifiées avant d'être intégrées. Cette année pour tenter de combler ses manques, nous avons ajouté FreeDict ⁸ qui est aussi sous licence GPL. Mais, quelle que soit la qualité et la complétude des dictionnaires employés, le problème de la traduction des bitermes reste entier : comme ils ne sont pas présents dans les dictionnaires, le seul moyen dont nous disposons pour engendrer leurs traductions est d'effectuer une traduction mot-à-mot, à partir de toutes les combinaisons possibles. Et l'inconvénient de cette méthode est qu'elle engendre beaucoup de bruit, puisqu'aucun des termes composant le biterme n'est désambiguïsé. Ainsi, en fusionnant les différentes traductions des deux dictionnaires pour les 200 questions nous obtenons 991 bitermes, soit une moyenne de presque 5 bitermes par question, alors qu'elle est de l'ordre de un biterme par question à l'issue de l'analyse des questions en français. Pour réduire ce bruit, une solution est de valider les bitermes en constatant leur présence dans la collection de documents. [GRE 99] décrit une expérience semblable en traduction automatique. Il utilise le Web afin d'ordonner des traductions de groupes nominaux, en particulier lorsqu'il s'agit de bitermes.

Nous avons donc utilisé Fastr ([JAC 96]), qui reconnaît dans un corpus des termes (mono ou multitermes) sous leur forme d'origine ou sous forme d'une variante sémantique ou syntaxique. Pour éviter d'appliquer Fastr à toute la collection, nous avons constitué un sous-ensemble de documents pertinents. Pour cela, nous construisons d'abord une requête booléenne avec chaque terme plein de la question, et les différentes traductions comme alternatives. Si nous n'obtenons pas suffisamment de documents (soit 1000 ici), nous réitérons l'interrogation en enlevant les termes les moins spécifiques et en gardant les noms propres et les nombres, le cas échéant. Si le nombre de documents cherché n'est toujours pas atteint, nous retenons les 1000 premiers documents obtenus par une interrogation vectorielle avec une requête contenant tous les termes. Le tableau 2 présente les résultats de l'application de Fastr sur ce sous-ensemble de documents. 251 bitermes ont été trouvés soit 25 %, ce qui nous ramène à une moyenne plus raisonnable de 1,26 bitermes par question. Parmi eux 30 % sont présents seulement à l'identique : il s'agit en général des noms propres. Enfin, 48,8 % de ces bitermes trouvés sont présents mais seulement sous forme de variante : bien souvent, il s'agit de bitermes pour lesquels un des termes est mal traduit. Par exemple, *oil importation* n'est pas présent tel quel dans la collection mais sous la forme *oil import*. Il serait donc pertinent de remplacer ces bitermes par leur variante la plus représentée dans la collection. Quand un biterme est trouvé dans la collection, il est validé, puis la traduction des termes qui le composent est elle-même validée et les autres traductions sont écartées. La validation des bitermes permet ainsi de valider les monoterme.

7. <http://magic-dic.homeunix.net>

8. <http://www.freedict.org>

Nombre de bitermes à valider	991
Bitermes trouvés dans la collection	251 - 25%
Bitermes trouvés sans variante	77 - 30%
Bitermes trouvés avec variantes seulement	120 - 48,8%

Tableau 2. Les bitermes de *Magic-Dic* et de *FreeDict* validés par *Fastr* sur un sous-ensemble de 189 967 documents.

	Questions en français	Questions traduites par Reverso	Questions en anglais (VO)
Termes	1180	1122	1163
Bitermes	272	204	261
Bitermes erronés	33	38	27
Bitermes en commun	-	106	

Tableau 3. Les bitermes dans les différents fichiers de questions

A l'issue de l'évaluation de CLEF 2005, nous disposons de la version originale des questions en anglais ⁹, et de la traduction anglaise automatique, beaucoup moins correcte, réalisée par Reverso ¹⁰. Pour chacun de ces ensembles de questions, comme le montre le tableau 3, le nombre de termes et de bitermes est assez proche. A la main, nous avons évalué combien de bitermes sont erronés, là encore les chiffres sont proches. En ce qui concerne les questions traduites automatiquement, il existe deux raisons principales pour lesquelles un biterme peut être erroné :

- dans deux tiers des cas, la constitution du biterme elle-même est fautive, les termes sont associés à tort ; c'est le cas par exemple pour *many country* dans la question *How many countries joined the international coalition to restore the democratic government in Haiti ?* ¹¹

- dans un tiers des cas, un des termes composant le biterme est mal traduit ou non traduit, comme *movement zapatiste* provenant de la question *What carry the courtiers of the movement zapatiste in Mexico ?* ¹²

Plus intéressant, nous avons pu calculer que parmi les 204 bitermes issus de la traduction Reverso, 106 sont également présents à l'identique dans la version originale anglaise. Parmi les 98 restants, 38 sont erronés (pour les raisons mentionnées ci-dessus), il reste donc 60 bitermes, ni erronés ni communs à la version originale. Quelques-uns sont formés de termes traduits différemment mais corrects néanmoins ;

9. C'est à partir de cette version originale en anglais que les questions sont traduites manuellement dans les différentes langues sources de la campagne CLEF.

10. <http://www.reverso.net/>

11. Qui est un exemple de très bonne traduction donnée par Reverso

12. Dont la bonne traduction serait *What do supporters of the Zapatistas in Mexico wear ?*

la plupart sont construits avec une syntaxe différente, ce qui n'est pas surprenant puisque la syntaxe des questions dont ils sont issus est bien souvent différente.

On peut conclure que si Reverso produit des questions mal formées du point de vue de la syntaxe, les termes qu'il utilise sont la plupart du temps bien choisis. On peut donc penser que, moyennant un programme de réalignement et en s'aidant également des informations apportées par le passage de Fastr, Reverso pourrait nous aider à déterminer les bitermes corrects parmi tous les bitermes traduits à l'*aveugle* à partir des traductions de Magic-Dic et de FreeDict.

6. Analyse des questions multilingue

Nous avons développé aux cours des évaluations, une analyse des questions bilingue pour notre système. Celle-ci s'appuie sur l'étiquetage morpho-syntaxique et l'analyse syntaxique des questions. Puis différentes informations sont déterminées à partir de ces deux analyses : reconnaissance du type attendu pour la réponse, de la catégorie de la question, des contextes temporels... Il existe bien entendu des lexiques et patrons spécifiques à chaque langue traitée, mais le coeur de ce module est indépendant de la langue. Il a été évalué sur des corpus de questions similaires en français et en anglais, et ses performances dans les deux langues sont très proches, de l'ordre de 90% de rappel et de précision pour la reconnaissance du type attendu de la réponse par exemple (pour plus de précisions, voir [LIG 06]).

Comme présenté ci-dessus, notre système s'appuie sur deux stratégies distinctes : l'analyse dans la langue d'origine, puis une traduction terme à terme ou la traduction de la question dans la langue cible, puis son analyse. L'inconvénient principal de cette seconde stratégie est que l'analyse de la question est fortement perturbée par toute erreur de traduction. Nous allons ici évaluer l'influence de ces erreurs de traduction sur les performances de notre système Reverso+QALC, et essayer de contourner cette difficulté, en tirant parti de la question dans la langue cible avant de la traduire. Si l'on cherche à évaluer cette seconde stratégie, il convient de se demander dans un premier temps quelle est l'influence d'une traduction un peu trop approximative sur les performances de notre système. Une erreur de traduction va se refléter à la fois sur les termes de la question, et sur l'analyse de la question, qui peut être perturbée par une construction anglaise peu réaliste. Ainsi, la question *Quelles sont les deux villes allemandes reliées à Hambourg par des trains express à grande vitesse ?* est traduite par *What are both German cities connected in Hamburg by express trains with big speed ?*, ce qui empêche notre système de l'analyser correctement.

Afin de contourner les problèmes de traduction, il est possible de travailler en amont ou en aval de celle-ci. [AHN 04] effectuent ce type de pré et post-corrections, grâce à l'utilisation de règles surfaciques de reformulation. Notre système actuel tient ainsi compte de quelques-unes des erreurs de traduction les plus fréquentes, en permettant au module d'analyse de la question de relâcher ses règles en cas de question traduite. Ainsi, une question de définition comme "Qu'est-ce que l'UNITA ?", traduite

“What UNITA ?”, au lieu de “What is the UNITA ?”, sera néanmoins bien considérée comme une demande de définition par nos patrons. Cependant, ce genre de correction dépend fortement des formes des questions à traiter, ainsi que des erreurs des traducteurs utilisés. Nous proposons donc un autre type de post-traitement, qui tire parti de l’aspect multilingue de la tâche, pour tenter d’améliorer l’analyse des questions traduites en tenant compte des possibilités d’erreur dans la traduction.

Notre solution pour tenter d’améliorer notre gestion des approximations dans la traduction consiste à effectuer une double analyse de la question, dans les langues source et cible, et de reporter les informations (ou tout du moins une partie de ces informations) retournées par l’analyse en langue source dans l’analyse en langue cible. Ceci est possible car d’une part, notre système traite les deux langues de façon parallèle, et d’autre part, un certain nombre des informations fournies par le module d’analyse des questions sont identiques en anglais et en français, comme par exemple la catégorie de la question ou le type attendu de la réponse si c’est une entité nommée (EN). Plus précisément, et dans le cas considéré ici d’un système partant de questions françaises pour analyser des documents en anglais, nous proposons d’effectuer l’analyse de la question en français, et de sa traduction en anglais, puis de reporter la catégorie de la question et le type EN attendu trouvés en français dans l’analyse de la question traduite en anglais. Ces informations trouvées sur la langue source devraient être plus fiables car issues d’une question bien formée. Par exemple, pour la question ‘*Combien de communautés Di Mambro a-t-il crée ?*’, la traduction est *How many Di Mambro communities has he create ?*, empêchant une analyse correcte. L’analyse de la question française est alors utilisée, et fournit la bonne catégorie *combien*, ainsi que le type d’entité nommée attendu *NUMBER*. Ces informations sont ensuite transférées dans l’analyse de la question anglaise, car elles seront plus fiables que celles calculées sur la question traduite. Ces caractéristiques sont utilisées à deux étapes de la chaîne de question-réponse principalement : lors de la sélection des phrases candidates, et lors de l’extraction de la réponse. L’amélioration de leur fiabilité devrait donc permettre d’augmenter le nombre de réponses correctes après ces deux étapes.

Pour tester cette stratégie, nous avons mené une expérience en nous fondant sur la tâche FR-EN de CLEF 2005, et ses 200 questions en français. Nous avons ainsi fait tourner notre système de question-réponse avec trois fichiers de questions différents :

- Le premier (appelé Fichier anglais par la suite) contenait toutes les questions traduites manuellement (par les organisateurs de CLEF) en anglais. Ce fichier servira de fichier test, car les performances de notre système sur ce fichier représentent celles qu’il pourrait atteindre sans problèmes de traduction.
- Le second (appelé Fichier traduit par la suite) contenait l’analyse des questions traduites en anglais, puis analysées, sans aucune modification.
- Le dernier (appelé Fichier amélioré par la suite) contenait l’analyse des questions traduites en anglais, puis analysées, mais dans lequel les catégories et types EN attendus avaient été remplacés par ceux de l’analyse des questions en français.

	Fic. traduit	Fic. amélioré	Fic. anglais
Nombre de questions EN	112	111	115
Phrases correctes aux 5 1ers rangs	46	48	57
Phrases correctes au 1er rang	24	24	34
Réponses correctes aux 5 1ers rangs	36	38	39
Réponses correctes au 1er rang	24	25	30
Nombre de questions non EN	88	89	85
Phrases correctes aux 5 1ers rangs	40	39	48
Phrases correctes au 1er rang	28	28	35
Réponses correctes aux 5 1ers rangs	16	17	20
Réponses correctes au 1er rang	10	11	11
Nombre de questions total	200	200	200
Phrases correctes aux 5 1ers rangs	86	87	105
Phrases correctes au 1er rang	52	52	69
Réponses correctes aux 5 1ers rangs	52	55	59
Réponses correctes au 1er rang	34	36	41

Tableau 4. Performances de notre système (en termes de nombre de réponses correctes sur 200 questions), pour chaque fichier d'entrée d'analyse de la question

Nous avons ensuite recherché le nombre de réponses correctes pour chaque fichier, après la sélection des phrases, et après l'extraction de la réponse. Le tableau 4 indique, pour chaque fichier, le nombre de questions attendant une entité nommée comme réponse, n'attendant pas une EN comme réponse, et le nombre total, ainsi que les résultats de notre système pour chaque type de questions. On y trouve le nombre de réponses correctes¹³ aux cinq premiers rangs, ainsi qu'au premier rang, d'abord pour les phrases réponses ("réponses longues") puis pour les réponses courtes.

Ces résultats nous montrent que le transfert d'informations de la langue source à la langue cible améliore significativement les résultats du système. Sur l'ensemble des questions, le nombre de bonnes réponses est passé de 34 sur le fichier de questions traduites à 36 sur le fichier amélioré pour le premier rang, et de 52 à 55 pour les 5 premiers rangs, les performances se rapprochant des résultats du système monolingue (41 réponses correctes au premier rang, et 59 dans les 5 premiers rangs). Pour la sélection des phrases, on peut noter que les performances varient peu entre le fichier de départ et le fichier amélioré.

Il est intéressant de remarquer que la différence entre systèmes monolingue et bilingue est nettement plus marquée après la sélection des phrases qu'après l'extraction de la réponse, ce qui tendrait à montrer que cette dernière étape est moins sensible aux erreurs de traduction. De plus, cette expérience montre que cette étape peut être

13. Les questions ont été testées avec des patrons de réponses provenant en partie des réponses des participants et en partie d'une recherche dans le corpus.

améliorée grâce à un transfert d'informations d'une langue à l'autre. Afin d'étendre cette stratégie, nous pourrions mettre en place un appariement mot à mot des termes de la question, afin de connaître, pour chaque terme français, son équivalent anglais, et de pouvoir traduire toutes les informations données par l'analyse française de la question. Ainsi, les erreurs de l'analyse des questions traduites seraient minimisées.

7. Conclusion

Notre système de question-réponse multilingue a pour originalité d'utiliser en parallèle les deux stratégies les plus couramment utilisées pour le passage d'une langue à une autre dans le domaine, ce qui nous permet de bénéficier des avantages de chacune. Cependant, chaque méthode présente des insuffisances, que nous nous sommes attachés dans cet article à évaluer et à prévenir.

Dans le cas de la traduction terme à terme, nous tirons parti de la présence de bitermes dans la question pour limiter les ambiguïtés possibles. En validant les bitermes qui sont présents dans la collection, nous améliorons la qualité de traduction des bitermes et des monoterms qui les composent. Cette amélioration devrait nous permettre d'effectuer une meilleure sélection des phrases candidates.

Dans le cas de la traduction de la question, nous utilisons les informations déduites de la question en langue source pour éviter les problèmes dus aux questions mal formées. Cette stratégie permet de résoudre, en partie tout du moins, la non grammaticalité des questions traduites ; l'appariement des termes des questions françaises et anglaises permettrait de transférer effectivement toutes les informations. En revanche les erreurs de désambiguïsation dues à la traduction demeurent.

8. Bibliographie

- [AHN 04] AHN K., ALEX B., BOS J., DELMAS T., LEIDNER J., SMILLIE M., « Cross-lingual Question Answering with QED », *Working Notes, CLEF Cross-Language Evaluation Forum*, Bath, UK, 2004, p. 335-342.
- [FER 02] FERRET O., GRAU B., HURAUPT-PLANTET M., ILLOUZ G., JACQUEMIN C., MONCEAUX L., ROBBA I., VILNAT A., « How NLP Can Improve Question Answering », *Knowledge Organization*, vol. 29, n° 3-4, 2002, p. 135-155.
- [GRA 05a] GRAU B., ILLOUZ G., MONCEAUX L., ROBBA I., VILNAT A., FERRET O., KATTEB F. E., « Application de plusieurs stratégies pour trouver des réponses en anglais à des questions posées en français », *Actes de la 8ème Conférence Internationale sur le Document Electronique*, Beyrouth, Liban, 2005.
- [GRA 05b] GRAU B., LIGOZAT A.-L., ROBBA I., SIALEU M., VILNAT A., « Term Translation Validation by Retrieving Bi-terms », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.
- [GRE 99] GREFENSTETTE G., « The World Wide Web as a resource for example-based machine translation tasks », *ASLIB Conference on Translating and the Computer*, vol. 21, London, UK, 1999.

- [HAR 05] HARTRUMPF S., « University of Hagen at QA@CLEF 2005 : Extending Knowledge and Deepening Linguistic Processing for Question Answering », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.
- [JAC 96] JACQUEMIN C., « A symbolic and surgical acquisition of terms through variation. », *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, , 1996, p. 425-438, Springer, Heidelberg.
- [JIJ 04] JIKKOUN V., MISHNE G., DE RIJKE M., SCHLOBACH S., AHN D., MULLER K., « The University of Amsterdam at QA@CLEF2004 », *Working Notes, CLEF Cross-Language Evaluation Forum*, Bath, UK, 2004, p. 321-325.
- [LAU 05] LAURENT D., SÉGUÉLA P., NÈGRE S., « Cross Lingual Question Answering using QRISTAL for CLEF 2005 », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.
- [LIG 06] LIGOZAT A.-L., GRAU B., ROBBA I., VILNAT A., « L'extraction des réponses dans un système de question-réponse », *Traitement Automatique des Langues Naturelles (TALN 2006)*, Leuven, Belgium, 2006.
- [NEU 04] NEUMANN G., SACALEANU B., « Experiments on Robust NL Question Interpretation and Multi-layered Document Annotation for a Cross-Language Question / Answering System », *Working Notes, CLEF Cross-Language Evaluation Forum*, Bath, UK, 2004, p. 311-320.
- [NEU 05] NEUMANN G., SACALEANU B., « DFKI's LT-lab at the CLEF 2005 Multiple Language Question Answering Track », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.
- [PAB 05] DE PABLO-SÁNCHEZ C., GONZÁLEZ-LEDESMA A., MARTÍNEZ-FERNÁNDEZ J. L., GUIRAO J. M., MARTINEZ P., MORENO A., « MIRACLE's 2005 Approach to Cross-Lingual Question Answering », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.
- [PER 04] PERRET L., « Question Answering System for the French Language », *Working Notes, CLEF Cross-Language Evaluation Forum*, Bath, UK, 2004, p. 295-305.
- [SUT 05] SUTCLIFFE R. F., MULCAHY M., GABBAY I., O'GORMAN A., WHITE K., SLATTERY D., « Cross-Language French-English Question Answering using the DLT System at CLEF 2005 », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.
- [TAN 04] TANEV H., NEGRI M., MAGNINI B., KOUYLEKOV M., « The DIOGENE Question Answering System at CLEF-2004 », *Working Notes, CLEF Cross-Language Evaluation Forum*, Bath UK, 2004, p. 325-333.
- [TAN 05] TANEV H., KOUYLEKOV M., MAGNINI B., NEGRI M., SIMOV K., « Exploiting Linguistic Indices and Syntactic Structures for Multilingual Question Answering : ITC-first at CLEF 2005 », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.
- [VAL 05] VALLIN A., MAGNINI B., GIAMPICCOLO D., AUNIMO L., AYACHE C., OSENOVA P., PEAS A., DE RIJKE M., SACALEANU B., SANTOS D., SUTCLIFFE R., « Overview of the CLEF 2005 Multilingual Question Answering Track », *Working Notes, CLEF Cross-Language Evaluation Forum*, Vienna, Austria, 2005.